Web-based Supplementary Materials

A Measurement Error Model for Heterogeneous Capture Probabilities in Mark-Recapture Experiments: An Estimating Equation Approach

Richard Huggins and Wen-Han Hwang

1 Web Appendix A: Robustness Analysis

The proposed approach depends on the conditional means and variances given by (3) and (4), which were derived for an approximating normal model. In order to examine the robustness of the estimators to this assumption, we conducted another simulation study where the latent covariates are not normal distributed. The design of this study is the same as that of in Section 6 except the covariates ω_i were generated from three different distributions: a mixed normal distribution (which has heavy tails), a chi-square distribution with 10 degrees of freedom (which has a positive skewness), and a transformation of this chi-square distribution from the random variable X to $-X$ (which has a negative skewness). All distributions were transformed to have mean 45 and variance 1. The mixture normal was derived from two normal distributions with means $(-2/\sqrt{5}, 2/\sqrt{5})$ and variances $(1/5, 1/5)$ and the mixture percentage was fixed at 0.5 so that the mixture distribution also has mean 0 and variance 1. The results are reported in Tables S1-S3. In these tables we also reported the estimated population size $\hat{N}_\omega$ and regression coefficient $\hat{\beta}_\omega$ that would have been computed if the ω_i were observable for captured individuals.

The long tailed mixture normal distribution in Table S1 gives results that are similar too but not as good as Table 2. There is evidence of bias in the estimation of β for all three procedures. However, the bias in the estimated population sizes is least for the EERRC. As expected, none of the procedures perform as well as when the ω_i were observable. Interestingly, in Table S2 the positively skewed distribution results in positive bias in the EERRC and the standard errors are over estimated. The negatively skewed distribution in Table S3 results in negative bias in estimating β and N, although again due to the skewed distribution of the population size estimator the median differs from the true population size more than does the mean. Due to the nature of the Horvitz-Thompson estimator this bias is not unexpected with a negatively skewed distribution for the covariates. Apart from the positively skewed case of Table S2 the EERRC performs the best and both it and the RRC outperform the naive method.

2 Web Appendix B: Data of Example

The capture history of *Prinia flaviventris* is listed in Table S4-S5. We show the identification number (ID), the collected time by weekly (Week), and the wing length in mm (Wing). Three records of wing length are missing and denoted by NA. The computations were carried out using the statistics programming interface R and the code can be obtained from the second author.

	$N = 600$				$\beta = 0.6$			
	$\hat{N}_\omega$	$\hat{N}_0$	$\hat{N}_R$	$\hat{N}_E$	$\hat{\beta}_\omega$	$\hat{\beta}_0$	$\hat{\beta}_R$	$\hat{\beta}_E$
$\sigma_\epsilon^2 = 0.4$								
AVE	637.6	560.8	590.3	607.0	0.577	0.355	0.485	0.534
MED	603.7	542.7	563.8	579.8	0.570	0.347	0.471	0.515
SD	167.1	109.7	132.0	142.7	0.215	0.142	0.203	0.228
A.SE	161.0	108.2	139.2	134.4	0.214	0.157	0.210	0.213
SDMAD	133.0	99.8	115.8	125.3	0.211	0.147	0.199	0.230
M.SE	130.6	96.3	115.5	114.1	0.210	0.154	0.206	0.202
RMSE	171.1	116.4	132.2	142.8	0.216	0.283	0.233	0.238
MAE	121.8	95.5	103.9	109.2	0.173	0.251	0.191	0.194
CP	0.936	0.833	0.897	0.893	0.950	0.628	0.911	0.915
$\sigma_\epsilon^2 = 0.8$								
AVE	642.0	537.8	587.2	595.0	0.587	0.274	0.484	0.555
MED	601.1	526.1	567.1	583.2	0.578	0.278	0.468	0.529
SD	180.9	92.5	138.6	127.8	0.217	0.113	0.245	0.288
A.SE	165.1	95.4	146.8	129.2	0.213	0.131	0.233	0.275
SDMAD	125.9	83.1	113.0	114.5	0.198	0.106	0.202	0.222
M.SE	130.3	89.3	121.7	120.1	0.210	0.130	0.211	0.230
RMSE	185.5	111.4	139.1	127.7	0.217	0.345	0.271	0.291
MAE	122.8	94.2	100.4	97.1	0.168	0.326	0.210	0.208
CP	0.927	0.786	0.889	0.879	0.939	0.279	0.863	0.909

Table S1: Results from 500 simulations with the covariate ω has a mixed normal distributed. The estimates $\hat{N}_0$ and $\hat{\beta}_0$ correspond to the naive method, $\hat{N}_R$ and $\hat{\beta}_R$ to the modified RRC method, and $\hat{N}_E$ and $\hat{\beta}_E$ to the EERRC. The estimator $\hat{N}_\omega$ and regression coefficient $\hat{\beta}_\omega$ is that would have been computed if the ω_i were observable for the captured individuals. We give the average (AVE), the median (Med), the sample standard deviations (SD), the rescaled median absolute deviation (RMAD), the average of estimated standard errors (A.SE), the median of estimated standard errors (M.SE), the sample root mean squared error (RMSE), the sample mean absolute error (MAE), and the coverage percentage (CP) of the nominal 95% confidence interval. Here $\bar{D} = 175, \overline{\sum Y_i} = 216$.

	$N = 600$				$\beta = 0.6$			
	$\hat{N}_\omega$	$\hat{N}_0$	$\hat{N}_R$	$\hat{N}_E$	$\hat{\beta}_\omega$	$\hat{\beta}_0$	$\hat{\beta}_R$	$\hat{\beta}_E$
$\sigma_\epsilon^2 = 0.4$								
AVE	616.7	617.9	642.0	645.6	0.599	0.541	0.624	0.667
MED	600.2	596.8	616.5	619.0	0.598	0.543	0.623	0.666
SD	122.5	129.9	139.7	143.2	0.087	0.091	0.104	0.123
A.SE	122.0	127.4	140.1	137.9	0.088	0.086	0.099	0.110
SDMAD	119.5	126.2	130.8	129.1	0.086	0.092	0.102	0.123
M.SE	114.7	117.9	127.7	127.2	0.086	0.084	0.096	0.105
RMSE	123.5	131.0	145.8	150.1	0.087	0.108	0.106	0.140
MAE	94.8	100.1	107.9	110.8	0.069	0.086	0.084	0.109
CP	0.946	0.946	0.960	0.958	0.962	0.872	0.956	0.898
$\sigma_\epsilon^2 = 0.8$								
AVE	620.3	608.2	667.4	662.2	0.599	0.482	0.648	0.714
MED	599.0	581.1	631.2	638.4	0.598	0.488	0.632	0.697
SD	128.7	134.2	161.3	151.8	0.091	0.093	0.123	0.156
A.SE	123.5	126.9	159.0	143.9	0.090	0.083	0.113	0.151
SDMAD	105.2	107.1	132.7	127.9	0.089	0.101	0.123	0.139
M.SE	112.5	113.4	139.4	133.2	0.088	0.083	0.108	0.130
RMSE	130.1	134.3	174.7	163.9	0.091	0.150	0.132	0.193
MAE	96.1	101.3	122.6	117.9	0.071	0.126	0.103	0.145
CP	0.934	0.930	0.956	0.956	0.942	0.673	0.918	0.906

Table S2: Results from 500 simulations with the covariate ω is chi-square distributed. Here $\bar{D} = 172, \overline{\sum Y_i} = 226$. The notation is as in Table S1.

	$\hat{N}_\omega$	$\hat{N}_0$	$\hat{N}_R$	$\hat{N}_E$	$\hat{\beta}_\omega$	$\hat{\beta}_0$	$\hat{\beta}_R$	$\hat{\beta}_E$
		$N = 600$				$\beta = 0.6$		
$\sigma_\epsilon^2 = 0.4$								
AVE	646.7	566.5	598.5	595.4	0.580	0.296	0.456	0.488
MED	605.6	550.7	564.5	574.0	0.560	0.284	0.423	0.478
SD	178.6	102.8	142.3	121.5	0.254	0.155	0.270	0.279
A.SE	182.4	105.7	150.0	119.9	0.282	0.187	0.288	0.267
SDMAD	133.9	95.5	110.6	108.3	0.252	0.159	0.240	0.245
M.SE	134.3	94.8	113.0	107.5	0.278	0.185	0.270	0.237
RMSE	184.5	108.0	142.2	121.4	0.255	0.341	0.306	0.301
MAE	124.9	89.3	104.6	95.6	0.205	0.307	0.246	0.231
CP	0.921	0.839	0.893	0.905	0.964	0.607	0.921	0.919
$\sigma_\epsilon^2 = 0.8$								
AVE	624.4	541.4	574.1	578.3	0.532	0.182	0.365	0.464
MED	579.5	529.7	550.9	556.9	0.518	0.177	0.351	0.421
SD	170.0	90.4	129.4	117.2	0.276	0.128	0.306	0.462
A.SE	163.5	90.5	139.2	119.0	0.276	0.151	0.603	20.492
SDMAD	119.0	89.8	106.8	103.3	0.276	0.123	0.259	0.284
M.SE	116.3	83.6	104.2	103.6	0.271	0.149	0.282	0.287
RMSE	171.6	107.6	131.8	119.1	0.284	0.437	0.385	0.481
MAE	118.7	90.0	101.0	95.9	0.231	0.419	0.321	0.318
CP	0.896	0.787	0.839	0.870	0.907	0.195	0.848	0.892

Table S3: Results from 500 simulations with the covariate ω is a transformed chi-square distributed. Here $\bar{D} = 173, \overline{\sum Y_i} = 209$. The notation is as in Table S1.

ID	Week	Wing	ID	Week	Wing	ID	Week	Wing
A001	5	46	A022	5	47	A047	1	47
A001	10	46	A023	6	46	A048	2	43
A001	12	46	A023	7	46	A049	4	45
A001	14	46	A024	14	48	A050	4	45
A001	15	46	A025	14	47	A051	5	44
A002	9	48	A026	17	47	A052	5	47
A003	10	49	A027	1	46	A053	5	44
A003	15	49	A028	5	44	A054	5	44
A004	11	46	A029	5	NA	A055	5	45
A005	1	46	A030	3	44	A056	5	44
A005	7	46	A030	15	43	A057	6	45
A005	10	46	A031	14	44	A057	11	45
A005	13	47	A031	15	46	A058	6	45
A006	1	46	A032	5	45	A059	6	46
A007	3	47	A033	10	46	A060	6	45
A007	12	47	A033	11	46	A061	6	45
A008	6	44	A033	15	46	A062	6	45
A009	15	45	A034	10	43	A062	12	46
A010	10	48	A035	15	46	A063	6	46
A010	16	47	A036	6	46	A064	6	48
A011	4	44	A036	17	45	A065	6	45
A012	2	44	A037	1	44	A066	7	46
A013	10	47	A037	4	45	A066	11	46
A013	11	47	A038	15	44	A066	13	46
A014	4	47	A039	11	45	A067	7	46
A014	8	47	A040	2	47	A068	7	45
A015	5	NA	A041	10	45	A069	7	43
A015	9	48	A041	12	47	A070	7	46
A015	10	47	A042	1	47	A071	7	46
A016	10	45	A043	1	45	A072	7	44
A017	6	45	A044	1	45	A073	7	43
A018	11	45	A045	1	47	A074	7	46
A019	11	46	A045	16	46	A075	7	45
A020	14	48	A046	1	46	A076	7	43
A021	8	43	A046	8	45	A077	7	46

Table S4: Captures of *Prinia flaviventris* collected weekly by the Hong Kong Bird Society in 1993 at Mai Po Bird Sanctuary. Continued on next Table.

ID	Week	Wing	ID	Week	Wing	ID	Week	Wing
A078	7	43	A107	10	43	A138	14	45
A079	8	46	A108	10	45	A139	14	47
A080	8	45	A109	10	47	A140	14	47
A081	8	43	A110	10	45	A140	15	46
A082	8	45	A111	10	43	A141	14	45
A083	8	46	A112	10	45	A142	14	45
A083	11	46	A113	10	45	A143	14	46
A083	12	46	A114	10	47	A144	14	45
A084	8	44	A115	10	44	A145	14	46
A085	8	45	A116	10	46	A145	17	47
A086	8	44	A117	10	45	A146	14	45
A087	8	43	A118	11	43	A147	14	47
A088	8	44	A119	11	46	A148	14	43
A089	8	45	A120	11	44	A149	14	46
A089	11	44	A121	11	43	A150	15	45
A090	8	43	A122	11	45	A151	15	46
A091	8	45	A122	17	45	A152	15	46
A092	8	45	A123	12	45	A153	15	43
A093	8	46	A124	12	45	A154	15	45
A094	9	45	A125	12	45	A155	15	45
A095	9	44	A125	15	46	A156	2	46
A095	10	45	A126	12	45	A157	13	45
A096	9	44	A127	12	45	A158	9	45
A097	9	44	A128	12	44	A159	6	45
A098	10	45	A129	12	45	A160	13	45
A099	10	47	A130	12	45	A161	6	46
A100	10	47	A130	13	45	A162	12	44
A100	12	NA	A130	16	44	A163	1	45
A101	10	46	A131	13	46	A163	3	45
A102	10	44	A132	13	46	A164	1	45
A103	10	45	A133	13	46	A165	14	45
A104	10	45	A134	13	45			
A104	13	46	A135	13	45			
A105	10	47	A136	13	46			
A106	10	44	A137	13	46			

Table S5: Continued from previous Table. Captures of *Prinia flaviventris* collected weekly by the Hong Kong Bird Society in 1993 at Mai Po Bird Sanctuary.